

Deep Learning-Inspired Image Quality Enhancement



Ruxin Wang

Faculty of Engineering and Information Technology

University of Technology Sydney

A thesis submitted for the degree of

Doctor of Philosophy

2017

To my loving parents
Jianxia Liu and Jigang Wang

Certificate of Original Authorship

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Ruxin Wang

Acknowledgements

I would like to sincerely thank everyone who has helped me to finish my doctoral studies.

First of all, I would like to express my sincere appreciation and deep gratitude to my supervisor **Prof. Dacheng Tao**. He has given me trust and freedom to pursue my research interests, and provided constructive suggestions to help me out of difficulties. I can always benefit and learn a lot from various detailed discussions with him, and be excited and energised by his amazing insight, unlimited patience, generous support, and constant encouragement. I feel very lucky to have had him as my supervisor.

I also wish to express my sincere appreciation to **Prof. Xinge You** who was my advisor when I was in Huazhong University of Science and Technology for master study. I am so grateful for his guidance and support. I would have not come to the research field of image processing and computer vision, and would not have met Prof. Dacheng Tao and come to UTS without him.

I would also like to give special thanks to my excellent collaborators: Prof. Jun Yu, Prof. Richang Hong, Dr. Kun Zeng, and Dr. Xiaoyan Li for their brilliant work and timely support. I have also been fortunate to work and have discussions with many other brilliant researchers: Dr Chang Xu, Dr Nannan Wang, Dr Fei Gao, Assistant Professor Chaohui Wang, Prof. Chen Gong, Dr Yong Luo, Dr Tao Liu, Prof. Kaibing Zhang, Dr Lianyang Ma, A/Prof Chenping Hou, Dr Xiao Liu, Dr Weilong Hou, A/Prof Bo Du, Prof Shengzheng Wang, A/Prof Tao Lei, A/Prof Wankou Yang, A/Prof Shigang Liu, A/Prof Xianye Ben, A/Prof Xiong Wang, and Dr Qiong Wang. I wish to

express appreciation to all of them for their support and kind companion.

I am so grateful to my colleagues and friends I met in Sydney: Mingming Gong, Tongliang Liu, Shaoli huang, Qiang Li, Zhibin Hong, Meng Fang, Changxing Ding, Maoying Qiao, Zhe Xu, Long Lan, Wei Bian, Tianyi Zhou, Jun Li, Chunyang Liu, Bozhong Liu, Zhiguo Long, Guodong Long, Jing Jiang, Huan Fu, Baosheng Yu, Zhe Chen, Zijing Chen, Xiyu Yu, Liu Liu, Yali Du, Guoliang Kang, Hao Xiong, Jiayan Qiu, Jiankang Deng, Chaoyue Wang, Dayong Tian, Jiang Bian, Bo Han, Kede Ma, Tianrong Rao, Lingxiang Wu, Sujuan Hou, Xiaoqing Yin, Peicheng Zhou, Mingjin Zhang, Tao Zhang, Haifeng Liu, Peng Hao, Haishuai Wang, and Xun Yang. All my friends here have provided strong support during both happy and stressful time. I am also incredibly grateful to my old friends: Chengbin Yan, Digang Wang, Xinwei Liu, Zhe Wang, Hao Ding, Junge Shen, Xin Wang, Ting Yuan, Haoran Lv, and Yuannan Zhao, who have accompanied me from my bachelor's degree to doctoral studies. I owe my deepest thanks to all of them!

Finally, I would like to express deep-felt gratitude to my family: my parents, my grandparents, my uncles and aunties, and my cousins, for their endless love, trust, encouragement, and full support throughout my studies and life.

I dedicate this thesis to them.

Abstract

Enhancing image quality is a classical image processing problem that has received plenty of attention over the past several decades. A high-quality image is always expected in various vision tasks, and degradations such as noise, low-resolution, and blur are required to be removed. While the conventional techniques for this task have achieved great progress, the recent top performer, deep models, can substantially and significantly boost performance compared with conventional ones. The advantages of deep learning which enables it to achieve such success are its high representational capacity and the strong nonlinearity of the models. In this thesis, we explore the development of advanced deep models for image quality enhancement by researching several fundamental issues with different motivations.

In particular, we are first motivated by a pivotal property of the human perceptual system that similar visual cues can stimulate the same neuron to induce similar neurological signals. However, image degradations can result in the fact that similar local structures in images exhibiting dissimilar observations. While the conventional neural networks do not consider this important property, we develop the (stacked) non-local auto-encoder which exploits self-similar information in natural images for enhancing the stability of signal propagation in the network. It is expected that similar structures should induce similar network propagation. This is achieved by constraining the difference between the hidden representations of non-local similar image blocks during training. By applying the proposed model to image restoration, we then develop a “collaborative stabilisation” step to further rectify forward propagation.

When applying deep models to image quality enhancement tasks, we are concerned about which factor, receptive field size or model depth, is more critical. To determine the answer, we focus on the single image super-resolution task, and propose a strategy based on dilated convolution to investigate how the two factors affect the performance. Our findings from exhaustive investigations suggest that single image super-resolution is more sensitive to the changes of receptive field size than to model depth variations, and that the model depth must be congruent with the receptive field size to produce improved performance. These findings inspire us to design a shallower architecture which can save computational and memory cost while preserving comparable effectiveness with respect to a much deeper architecture.

Finally, we study the general non-blind image deconvolution problem. It is observed in practice that by using existing deconvolution techniques, the residual between the sharp image and the estimation is highly dependent on both the sharp image and the noise. These techniques require the construction of different restoration models for different blur kernels and noises, inducing low computational efficiency or highly redundant model parameters. Thus, for general purposes, we propose a method by designing a very deep convolutional neural network which can handle different kernels and noises, while preserving high effectiveness and efficiency. Instead of directly outputting the deconvolved results, the model predicts the residual between a pre-deconvolved image and the corresponding sharp image, which can make the training easier and obtain restored images with suppressed artifacts.

Contents

Contents	i
List of Figures	v
List of Tables	ix
1 Introduction	1
1.1 Sources of Image Quality Degradations	2
1.2 Image Formulation and Restoration	4
1.3 Philosophy of Image Quality Enhancement	8
1.4 Motivations	9
1.5 Summary of Contributions	10
2 Literature Review	13
2.1 Image Denoising	13
2.1.1 Spatial domain-based approaches	14
2.1.2 Transform domain-based approaches	15
2.1.3 Learning-based approaches	17
2.2 Image Super-resolution	18
2.2.1 Reconstruction-based super-resolution	18
2.2.2 Learning-based super-resolution	19
2.3 Image Deblurring	21
2.3.1 Methodology on deblurring	22
2.3.2 Modelling of spatially variant blur	25
2.4 Deep Learning	25

CONTENTS

2.4.1	Auto-encoder	26
2.4.2	Convolutional neural network	28
3	Non-local Auto-encoder with Collaborative Stabilization	31
3.1	Introduction	32
3.2	Non-local auto-encoder	34
3.2.1	Turbulence of an auto-encoder	35
3.2.2	Modelling the non-local auto-encoder	37
3.2.3	Stacked non-local auto-encoders	40
3.2.4	Collaborative stabilization	41
3.3	Relationships between non-local auto-encoder and the literature .	44
3.4	Realizations in training and testing	46
3.4.1	Training strategies	46
3.4.2	Testing strategies	49
3.5	Experiments	50
3.5.1	Learned weights and model stability	50
3.5.2	Image denoising	52
3.5.3	Image super-resolution	54
3.6	Conclusions	61
4	Receptive Field Size <i>vs.</i> Model Depth	65
4.1	Introduction	66
4.2	Related Work on Dilated Convolution	68
4.3	Basic Settings for Comparison	69
4.4	Effects of Receptive Field Size on SISR	72
4.4.1	L10 CNN models	74
4.4.2	L6 and L20 CNN models	77
4.4.3	An instance-level investigation	81
4.5	Effects of Model Depth on SISR	83
4.5.1	Insufficient receptive field size	85
4.5.2	Sufficient receptive field size	86
4.6	Discussion	91
4.6.1	Comparison with state-of-the-art	94

CONTENTS

4.7	Conclusion	96
5	Residual Learning in Non-Blind Image Deconvolution	99
5.1	Introduction	100
5.2	Related Work on Residual Learning	102
5.3	Deep CNNs for General Non-blind Deconvolution	104
5.3.1	Residual analysis	104
5.3.2	Network architecture	106
5.3.3	Training	107
5.3.4	Discussion	110
5.4	Experiments	112
5.4.1	Effects of residual learning	114
5.4.2	Evaluation on different depths	115
5.4.3	Comparison to the state-of-the-art	118
5.5	Conclusion	126
6	Conclusions	127
	References	131

CONTENTS

List of Figures

1.1	Examples of photography. (a) The image we want to see. (b)(c)(d) The images we may capture.	3
1.2	Examples of remote sensing and medical imaging. (a)(c) The images that are expected. (b)(d) The images that are degraded. . .	4
2.1	A typical architecture of the auto-encoder.	27
3.1	A stacked auto-encoders (in (a)) and its layer-wise responses (in (c)) with respect to the eight similar patches (in (b)). In each subfigure of (c), the x-axis indicates the indexes of neurons in the corresponding hidden layer, while the y-axis indicates the output values given those patches.	36
3.2	Statistics of layer-wise turbulences in stacked auto-encoders. For one of the five hidden layers, we denote by h_j the maximal response of the j -th neuron w.r.t. a set of similar patches, and by h_{max} the maximal response over all neurons in this layer. We identify neurons as activated if $h_j > 0.1 \times h_{max}$. Assuming the set of activated neurons to be $\{act\}$ with cardinality n and $p = 2$, the average turbulence is $\frac{\sqrt{\sum_{j \in \{act\}} (h_{0j}^2 - h_{1j}^2)}}{n}$, where the subscripts 0 and 1 indicate different instances. The standard derivation is calculated accordingly.	37
3.3	Non-local auto-encoder. \sim indicates that the two arguments are similar, while \simeq means that \mathbf{h} and \mathbf{h}_i are collaboratively similar. .	38
3.4	Manifold interpretation of the non-local auto-encoder.	41
3.5	Distribution of the neuron-wise turbulence.	43

LIST OF FIGURES

3.6	Randomly selected weights of the stacked non-local auto-encoders. (a) and (b) are from the first and last layers in image denoising task. (c) and (d) are from the first and last layers in image super-resolution task.	50
3.7	Statistics of layer-wise turbulences in stacked non-local auto-encoders.	51
3.8	PSNR(dB) and SSIM results of different image denoising methods on <i>lenna</i> ($\sigma = 25$).	55
3.9	PSNR(dB) and SSIM results of different image denoising methods on <i>house</i> ($\sigma = 50$).	56
3.10	Performances of SNA and SNA _{cs} (both of which are trained for $\sigma = 25$) with respect to different noise levels.	57
3.11	Image super-resolution results of different methods when the up-scale factor is 2.	60
3.12	Image super-resolution results of different methods when the up-scale factor is 4.	61
3.13	Super-resolution performances of SNA and SNA _{cs} in handling noisy low-resolution images.	62
4.1	2D dilated filters. The solid circles indicate the filter parameters, while the hollow circles indicate zeros that are inserted during dilation.	68
4.2	Basic CNN architecture. The CNN model with dilation will be specified in each individual task.	70
4.3	The curves of test performances on Set5 for $\times 2$, $\times 3$, and $\times 4$ tasks.	72
4.4	The SISR performances of the L10 models with different number of dilated convolutional layers. The up-scaling factor used in each comparison is marked in the top of the subfigures. In each dataset, the positive value on the right of each bar indicates the improved quality by the corresponding model <i>w.r.t.</i> Bicubic interpolation. The minimum at x-axis is adjusted for better visualisation of the differences between the performances.	75
4.5	The SISR performances of the L6 models with different number of dilated convolutional layers.	79

LIST OF FIGURES

4.6	The SISR performances of the L20 models with different number of dilated convolutional layers.	80
4.7	Examples of local entropy.	82
4.8	Improved PSNR $\Delta_{L10D0}\mathcal{P}_{\text{best}}^s$ <i>v.s.</i> average local entropy for each image in BSD100 and Urban100. In each case, the best model (L10D2 for $\times 2$, L10D2 for $\times 3$, and L10D3 for $\times 4$) is used for evaluation.	83
4.9	SR examples of L10D2 for $\times 3$. From top to bottom: original HR image, restored image, residual image, and local entropy image. The left two columns are the examples that achieve the highest $\Delta_{L10D0}\mathcal{P}_{L10D2}^3$ in BSD100 and Urban100, respectively. The right two columns achieve the lowest $\Delta_{L10D0}\mathcal{P}_{L10D2}^3$ in the two datasets.	84
4.10	The SISR performances of the <i>LmD0Po</i> models for the investigation of model depth under the setting of an insufficient receptive field size.	87
4.11	The SISR performances of the <i>L6DnPo</i> models for the investigation of model depth under the setting of a sufficient receptive field size.	89
4.12	The SISR performances of the <i>L10DnPo</i> models for the investigation of model depth under the setting of a sufficient receptive field size.	90
4.13	SR examples with an up-scaling factor of 3. PSNR/SSIM values are marked under each subfigure.	96
4.14	SR examples with an up-scaling factor of 3 (<i>cont.</i>). PSNR/SSIM values are marked under each subfigure.	97
4.15	SR examples with an up-scaling factor of 3 (<i>cont.</i>). PSNR/SSIM values are marked under each subfigure.	98
5.1	Illustration of the image residual. (a) The original sharp image x . (b) The blur kernel k . (c) The blur image y . (d) The residual r_b . (e) The pre-deconvolved image \hat{x} . (f) The residual r	103
5.2	The network architecture. The model takes a pre-deconvolved image as input and estimates the residual image. $c = 3$ for colour images and $c = 1$ for grey images.	106

LIST OF FIGURES

5.3	Examples of training kernels.	107
5.4	The curve of validation performance during training.	110
5.5	Histograms of the original image and the residual image. From left to right: original sharp image, absolute residual image, and histograms.	111
5.6	The ratio of the histogram entropies between the original images and the corresponding absolute residual images.	112
5.7	Different mappings between the input space (the left of the arrows) and the output space (the right of the arrows). (a) An injective mapping is assumed between the input and output spaces. (b) In the residual learning case, the output space is shrunk by computing the residual. (c) In the denoising auto-encoder case, the input is corrupted by random noises to generate new input samples.	113
5.8	Test kernels including 8 motion kernels, 3 Gaussian kernels, 2 square kernels, and 2 disk kernels.	114
5.9	Comparisons between the models trained with and without residual learning.	115
5.10	Examples restored by DEBCNN_res and DEBCNN_nores. (It is better viewed by zoom-in.)	116
5.11	Performances of the models with different depths.	117
5.12	Visual examples (MotionA). It is better viewed by zoom-in.	122
5.13	Visual examples (GaussianA). It is better viewed by zoom-in.	123
5.14	Visual examples (SquareA). It is better viewed by zoom-in.	124
5.15	Visual examples (DiskB). It is better viewed by zoom-in.	125

List of Tables

3.1	Parameter settings in training	49
3.2	PSNR(dB) and SSIM results of image denoising ($\sigma = 25$). The best performance among the competitors (except DnCNN) is marked in bold.	53
3.3	PSNR(dB) and SSIM results of image denoising ($\sigma = 50$). The best performance among the competitors (except DnCNN) is marked in bold.	53
3.4	Effect of non-local regularizer on image denoising. Average performances are reported.	54
3.5	Running time comparison between different methods on image denoising. The methods with GPU implementations are marked with “*”.	58
3.6	PSNR(dB) and SSIM results of image super-resolution on <i>Set5</i> . .	58
3.7	PSNR(dB) and SSIM results of image super-resolution on <i>Set14</i> . .	59
3.8	Effect of non-local regularizer on image super-resolution. Average performances are reported.	62
3.9	Running time comparison between different methods on image SR with scale factor 2. The methods with GPU implementations are marked with “*”.	63
4.1	Training parameters.	71
4.2	Receptive field sizes of different dilated filters (3×3 -basic size). .	73
4.3	Architecture settings of the L10 CNN models.	74
4.4	Comparison of improved quality $\Delta_{L10D0}\mathcal{P}_{\text{best}}^s$ for the L10 models. .	77
4.5	Architecture settings of the L6 and L20 CNN models.	77

LIST OF TABLES

4.6	Comparison of improved quality $\Delta_{L6D0}\mathcal{P}_{\text{best}}^s$ for the L6 models. . .	81
4.7	Comparison of improved quality $\Delta_{L20D0}\mathcal{P}_{\text{best}}^s$ for the L20 models. .	81
4.8	Architecture settings of the <i>LmD0Po</i> CNN models.	86
4.9	Architecture settings of the <i>L6DnPo</i> CNN models.	88
4.10	Architecture settings of the <i>L10DnPo</i> CNN models.	88
4.11	Average PSNRs (dB) ($\times 2$) using different receptive field sizes (RFS) and model depths. The top 30% values are marked by bold. The top three values are marked by red, green, and blue, respectively.	92
4.12	Average PSNRs (dB) ($\times 3$) using different receptive field sizes (RFS) and model depths.	92
4.13	Average PSNRs (dB) ($\times 4$) using different receptive field sizes (RFS) and model depths.	93
4.14	Average PSNR(dB)/SSIM of different methods on Set5, Set14, BSD100, and Urban100.	95
5.1	Training parameters.	108
5.2	Performance comparison on BSD100. Average PSNR/SSIM are provided with the best value marked in bold. “-” indicates that the model is not suitable for the corresponding case.	119
5.3	Performance comparison on Set16. Average PSNR/SSIM are provided with the best value marked in bold. “-” indicates that the model is not suitable for the corresponding case.	119
5.4	Performance comparison on Set30. Average PSNR/SSIM are provided with the best value marked in bold. “-” indicates that the model is not suitable for the corresponding case.	120
5.5	Comparison of running time between different methods.	126